ESTATÍSTICA EXPLORATÓRIA

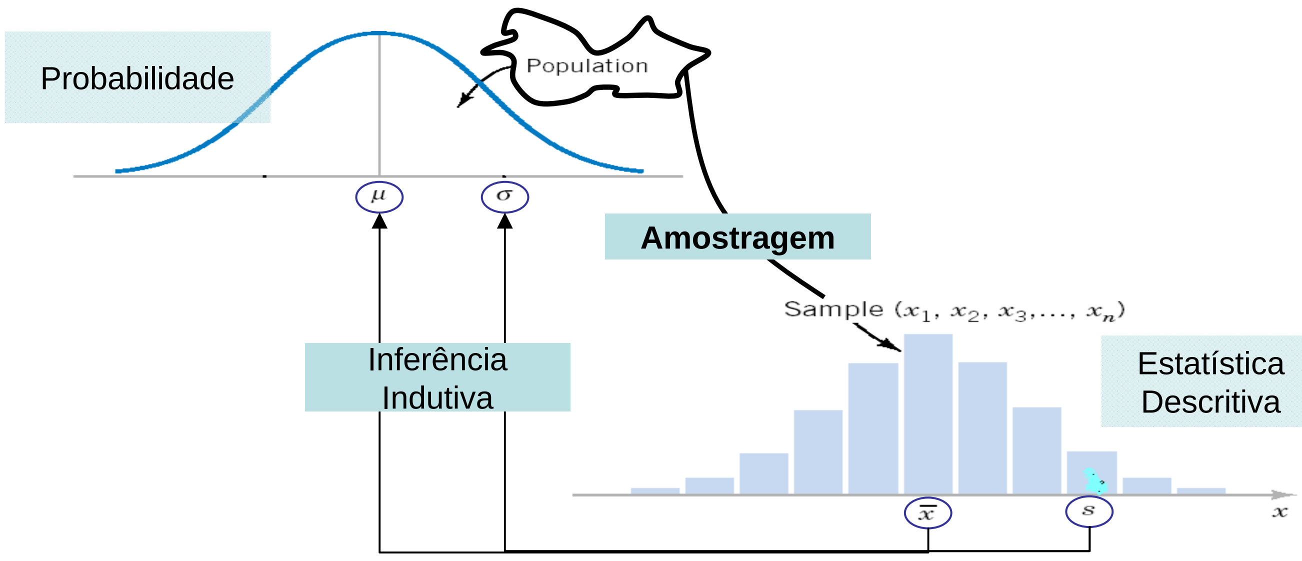
Prof Paulo Renato A. Firmino

praf62@gmail.com

Aulas 19-20

Inferência Indutiva - Definições

- Conceitos importantes
  - **Parâmetro**: função diretamente associada à população
    - É um valor fixo, mas desconhecido
  - **Estimador**: função diretamente associada a uma amostra aleatória
    - É uma variável aleatória
  - **Estimativa**: função diretamente associada a uma amostra observada
    - É um valor fixo e conhecido



Inferência Indutiva – Estimadores
Média & Teorema Central do Limite (TCL)

- Quando o tamanho da amostra (n) cresce, $\overline{X} \sim N(\mu, \sigma^2/n)$

  - Supõe-se que a amostra aleatória ($X_1, X_2, ..., X_n$) é tal que
    - $X_1, X_2, ..., X_n$ são VAs independentes entre si
    - $E(X_i) = \mu$
    - $V(X_i) = \sigma^2$

  - Se $X_i \sim$ Binomial (n=1, p), então $\hat{p} \sim N(p, p \cdot (1-p)/n)$
    - $E(X_i) = p$
    - $V(X_i) = p \cdot (1-p)$

Inferência Indutiva – Estimadores
Média: Distribuição t-Student

- Se σ^2 é desconhecida, $T = \dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$

    - S^2 é o estimador da variância populacional σ^2

Inferência Indutiva – Estimadores Média & TCL: Tamanho da Amostra

- Como estimar μ de maneira racional?
  - Suponha que deseja-se definir um n para o qual a probabilidade de $\overline{X}$ divergir (diferir) de μ em, no máximo, ε seja igual a $1 - \alpha$:

$$P\left(\left|\overline{X}-\mu\right| \le \varepsilon\right) = 1-\alpha = P\left(\frac{\left|\overline{X}-\mu\right|}{\sigma/\sqrt{n}} \le \frac{\varepsilon}{\sigma/\sqrt{n}}\right) =$$

$$= P\left(\left|Z\right| \le \frac{\varepsilon}{\sigma/\sqrt{n}}\right) \Rightarrow P\left(Z \le -\frac{\varepsilon}{\sigma/\sqrt{n}}\right) = \frac{\alpha}{2} \Rightarrow$$

$$\Rightarrow z_{\alpha/2} = -\frac{\varepsilon}{\sigma/\sqrt{n}} \Rightarrow n = \left(\frac{z_{\alpha/2}\cdot\sigma}{\varepsilon}\right)^2$$

  - Pode-se calcular n a partir da última igualdade acima
    - Supõe-se que $\overline{X}$ segue uma distribuição Normal, com média μ e variância σ^2/n
    - Caso σ^2 seja desconhecido, usa-se uma estimativa provinda de uma ***amostra*** (s^2) e considera-se uma **distribuição t-Student**

Inferência Indutiva – Estimadores
Média & TCL: Exercício 5

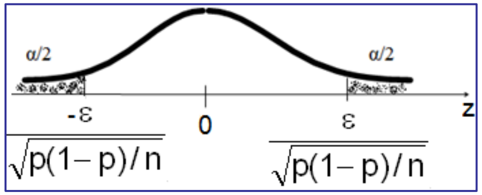
1. Deseja-se estimar o custo médio da cesta básica em dado país. Supondo que o desvio-padrão relacionado seja de R$50.00, qual deve ser o tamanho de uma amostra de cidades (n) para o qual a probabilidade de o custo médio da cesta básica amostral divergir do custo médio da cesta básica do país em, no máximo, R$110.00 seja igual a 95%?
2. Considere que a variância nominal definida para o tempo de vida de uma bateria (X) seja $\sigma^2 = 10000$ horas^2. Qual deve ser o tamanho de uma amostra, n, para o qual a probabilidade de o tempo médio de vida amostral divergir do tempo médio de vida da população de baterias em horas (μ) em, no máximo, 100 horas seja igual a 90%?
3. A previsão para a temperatura média em dado dia é baseada em uma amostra envolvendo dias semelhantes. Considerando um desvio-padrão de $\sigma = 4$ºC, divulga-se que a temperatura média estará em torno de 28ºC, com um erro de 2ºC para mais ou para menos. Divulga-se ainda que a probabilidade desta margem de erro não ser respeitada é de 1%. Qual tamanho amostral teria embasado esta estimativa?
4. Ainda sobre o problema da previsão do tempo (sob $\sigma = 4$ºC e um erro de 2ºC), baseando-se em uma amostra envolvendo 200 dias semelhantes, divulga-se que a temperatura estará em torno de 28ºC. Qual é a probabilidade de o erro máximo pré-estabelecido não ser respeitado?
5. Quais suposições embasam suas análises?

Inferência Indutiva – Estimadores
Média & TCL: Tamanho da Amostra

- E quando quer-se estimar p de maneira racional?
  - Deseja-se definir um n para o qual a probabilidade de $\hat{p}$ divergir (diferir) de p em, no máximo, ε seja igual a 1 – α:

$$P\left(|\hat{p}-p|\leq\varepsilon\right)=1-\alpha=P\left(\frac{|\hat{p}-p|}{\sqrt{p(1-p)/n}}\leq\frac{\varepsilon}{\sqrt{p(1-p)/n}}\right)=$$

$$=P\left(|Z|\leq\frac{\varepsilon}{\sqrt{p(1-p)/n}}\right)\Rightarrow$$



$$\Rightarrow P\left(Z\leq\frac{\varepsilon}{\sqrt{p(1-p)/n}}\right)=1-\frac{\alpha}{2}\Rightarrow z_{1-\alpha/2}=\frac{\varepsilon}{\sqrt{p(1-p)/n}}\Rightarrow n=\left(\frac{z_{1-\alpha/2}\cdot\sqrt{p(1-p)}}{\varepsilon}\right)^2$$

  - Pode-se calcular n a partir da última igualdade acima
    - Supõe-se que $\hat{p}$ segue uma distribuição Normal, com média p e variância $\sigma^2 = p(1-p)/n$
    - Como a variância é desconhecida, pode-se definí-la como a maior possível:

$$p(1\text{-}p)\leq 1/4\rightarrow n=\left(\frac{z_{1-\alpha/2}}{2\cdot\varepsilon}\right)^2$$

Inferência Indutiva – Estimadores
Média & TCL: Exercício 1

1. Qual deve ser o tamanho amostral, n, para o qual a probabilidade de a proporção amostral de votantes em B diferir (divergir) desta proporção nas eleições em, no máximo, 2 pontos percentuais (2%) seja igual a 90%?
2. Um negociante deseja conhecer o percentual de pessoas inadimplentes de dada localidade para planejar novos negócios. Quantas pessoas ele deve selecionar aleatoriamente de maneira que a probabilidade de a proporção amostral de inadimplentes divergir desta proporção em toda a região em, no máximo, 3 pontos percentuais (3%) seja igual a 95%?
3. Deseja-se determinar o tamanho amostral que garanta um erro percentual máximo de 1 ponto na diferença entre a proporção amostral e a populacional de pessoas que consideram justo terem sido deslocadas de seus lares em nome do desenvolvimento (instalação de hidrelétricas, preservação de reservas naturais, construção de rodovias, etc). Qual deve ser o tamanho da amostra para que a probabilidade do erro ser respeitado seja de: (a) 80%, (b) 90%, (c) 99%, (d) 99.9%?
4. Um telejornal anuncia que a proporção de brasileiros que aprova o governo do (a) presidente é de 40%, com uma margem de erro de 2 pontos percentuais para mais ou para menos. Ele acrescenta que 2000 pessoas de todo o país foram entrevistadas. Neste caso, qual é a probabilidade de que o erro percentual seja respeitado?
5. Quais suposições embasam suas análises?